ESTHER DYSON'S MONTHLY REPORT

Release 1.0®

VOLUME 20, NO. 10 | 25 NOVEMBER 2002 | www.edventure.com

Tapping into Grid Computing

INSIDE

GRID COMPUTING 1
Mapping the World of Grids 2
Grids, clusters, P2P, oh my!
Resource management
From Science to Commerce 6
Academic roots
Grids meet Web services
Standards and politics
Big Gridders 8
IBM
HP
Sun
Grid Specialists 10
Platform Computing
Avaki
Box: Entropia
The Mind Electric
Making the Grid 17
Power to the people
Complexity and decentralization
Box: Consumer grids
Grids and service business models
Resources & Contact Information 21

PC Forum, March 23-25, 2003.
See pages 22-23 for details!

BY KEVIN WERBACH

In the context of Release 1.0, *Kevin Werbach truly needs no introduction! He has been writing for this newsletter for four years, since joining us from the FCC where he was counsel for new technology policy. He has consistently championed decentralization, bottom-up architectures (whether for technology or policy), and peer-to-peer notions of all kinds. He has chronicled the not-yet-completed rise of Web services and other flat rather than hierarchical approaches to software, always with keen analysis of the policy as well as technical implications. Open spectrum, syndication, Weblogs and postmodern knowledge management are among the topics he brought to light in these pages.*

And now, with a new baby and a new house for his family, he's moving on in his career, becoming an occasional contributor to Release 1.0 *even as he works as an independent analyst and writes for other publications. He also has his own blog (http://werblog.com), of course!*

In the current issue, he outlines the decentralized underpinnings that will support Web services in particular and the sheer volume of organizational computing requirements in general. Just as we have a communication cloud – plug in anywhere – so will we have a computing cloud. Note that this is at the infrastructure/operating system level of fairly fungible computation and other resources, rather than at the application/function level where Web services will also some day offer piped-in services. (We'll be covering Web services next month, in an issue written by John Hagel of of John Hagel & Associates.)

We wish Kevin well in his new life, and we're looking forward to seeing him back in our pages from time to time! – Esther Dyson

{ *continued on page 2* }

Release 1.0® (ISSN 1047-935X) is published monthly except for a combined July/August issue by EDventure Holdings Inc., 104 Fifth Avenue, New York, NY 10011-6987; 1 (212) 924-8800; fax, 1 (212) 924-0240; www.edventure.com. It covers software, the Internet, e-commerce, convergence, online services, Web applications, technology policy, groupware, data networking, streaming media, enterprise applications, wireless communications, intellectual property and other unpredictable topics.

EDITOR-IN-CHIEF: Esther Dyson (edyson@edventure.com)

PUBLISHER: Daphne Kis (daphne@edventure.com)

CONTRIBUTING EDITOR: Kevin Werbach (kevin@edventure.com)

SYSTEMS DIRECTOR: Scott Giering (scott@edventure.com)

CIRCULATION MANAGER: Natasha Felshman (natasha@edventure.com)

ASSISTANT EDITOR: Christina Koukkos (christina@edventure.com)

CONSULTING EDITOR: Bill Kutik (bill@kutik.com)

Copyright © 2002, EDventure Holdings Inc. All rights reserved. No material in this publication may be reproduced without written permission; however, we gladly arrange for reprints, bulk orders or site licenses. Subscriptions cost $795 per year in the US, Canada and Mexico; $850 overseas.

What if any computing resource – processing, storage capacity, data, applications – could be an input to a larger process, regardless of where that resource was located? Such a vision is now thinkable, albeit not fully realized. Thanks to grid computing, we can ponder the potential of the universal computer... and can take advantage of pools of computing resources today for real applications.

Grid computing is moving from the academic to the commercial realm under the influence of Web services. As conceived, Web services use Internet-based protocols and programming models for distributed component applications. As implemented, they will also need to manage heterogeneous computing resources as a single entity, with all the necessary reliability, security, management and monitoring capabilities.

That's where grid computing comes in. Scientific researchers are finding common ground with enterprise IT architects, laying the groundwork for business and consumer applications of distributed computing. Grids are the low-level infrastructure for the emerging service-oriented, on-demand computing paradigm. In effect, grid software is a new kind of middleware – middleware that makes devices, and pools of devices, manageable as single entities. IBM's grid computing gm Tom Hawk jokingly calls it "griddleware."

Mapping the World of Grids

The influence of grids extends beyond today's relatively small commercial market. Research firm Grid Technology Partners pegs the current market at $180 million and, predictably, sees it growing to $4.1 billion in 2005. Sun, using a broader definition, estimates that the sector generated $1.5 billion in 2001 and will exceed $4 billion in 2003, still a tiny percentage of enterprise IT spending.

All the major IT vendors think grids will be a large opportunity. IBM has pledged to invest $10 billion in utility computing, with grid technologies being a big part of its vision. Paying customers are few and far between today, but almost every large enterprise with

high-end computing requirements is looking at grids and waiting for solutions to mature. A January 2002 survey by London-based Silico Research of senior executives at life sciences companies found that 69 percent use Linux clusters today, 46 percent use distributed computing to split tasks across machines and 23 percent have tried full-scale grid computing.

Plug me in

Grid computing draws on an analogy to the electricity distribution system. Power from any generating station can serve any customer who needs it, across a huge geographic area. Businesses and residential users need not worry about the intricacies of transformers and peak load management; that all happens magically through the power grid. They simply plug into a wall jack and pay a bill.

Grids bring some of that fluidity to computing. Agribusiness giant Monsanto develops new hybrid seeds through genomic analysis. It was crunching so much data that it maxed out the capacity of its largest mainframe. Rather than spend millions on a bigger machine, it used grid software (from Platform Computing, covered below) to allocate some of the processing to an under-utilized computing cluster at another location. The grid setup also allows Monsanto scientists to submit jobs directly from their desktops to the virtual supercomputer, with appropriate load balancing. All in all, the company avoided 90 percent of the expected hardware upgrade costs and can now process 20 to 50 percent more data per day.

The utility model doesn't apply perfectly to computing services, which are not as fungible as power, but it works as a high-level vision. The messy practical details bring out an interesting set of challenges around resource management, identity, and collaborative business. In fact, dealing with the messy details is the main problem. The issue is virtualization – making multiple resources appear as a single resource. The companies below are all developing software, not hardware, to manage heterogeneous hardware resources and present them as a single service.

With overall enterprise server utilization rates around 15 to 25 percent, grids hold the potential for major efficiency gains and cost avoidance. The early adopters in applying grid computing range from the serious (life sciences and financial services) to the fun (online gaming). IBM, HP and Sun are venturing into the field in a big way. Standards efforts are moving into high gear. Grids are going mainstream.

Grids, clusters, P2P, oh my!

This issue is the third installment in a trilogy that began with our reports on peer-to-peer (P2P) (SEE **RELEASE 1.0**, NOVEMBER-DECEMBER 2000) and Web services (SEE **RELEASE 1.0**, SEPTEMBER-OCTOBER 2001), both of which in turn linked back to our earlier coverage of component software, object orientation, clustering and parallel computing. All these involve the confluence of distributed computing technology with a ubiquitous and standards-based networking platform (the Internet). The profusion of concepts can easily create confusion. Marketing-driven labels and an unwillingness to acknowledge debts to earlier research work compound the problem.

So what are grids? "The Grid just means a lot of infrastructure doing cool stuff that is not located on either the client side or the server side," ventures The Mind Electric ceo Graham Glass (SEE PAGE 15). If that definition isn't precise enough, we offer the following taxonomy:

- **P2P:** networking devices, applications and data together horizontally through direct connections, rather than up and down though central servers. It's an architecture, not a business segment or a development model.
- **Web services**: a set of standard XML protocols for component-based applications. The resulting business models and methodologies emphasize services with many constituent parts, rather than atomic applications, ultimately operating across corporate boundaries.
- **Clustering**: operating a pool of similar computers as a single machine to handle a defined task.
- **Grid computing:** treating heterogeneous collections of computing resources as a single multi-tasking computer. Grid computing is broader than clustering because of its heterogeneity.

There are various kinds of computing grids, with varying capabilities. Astrophysics research doesn't have the same latency requirements as online gaming, or the same security issues as financial risk management. Some grids operate within a single company or research institution, while others operate across many, but all of them cross borders between physical machines and among other resources.[1]

---

[1] Though we focus on grids in this report, we should note that there are exciting developments in clustering as well. Open up a high-end Sun Fire 15K server from Sun, for example, and you'll find up to 106 UltraSparc microprocessors. The server's operating system parcels out work to those processors, allowing them to operate in parallel and deliver higher performance. Clusters do the same thing for physically separate machines, which can dramatically lower costs when those are commodity machines using Intel x86 chips and Linux software. The implications of clustering are less far-reaching than those of grids, which means its benefits may be realized more quickly.

One aspect of grid computing is what we called distributed computing in part II of our P2P report (SEE **RELEASE 1.0**, DECEMBER 2000). That approach focuses on harvesting spare cycles from Internet-connected PCs or enterprise workstations. Many of the cycle-sharing startups failed, and others have moved towards narrower applications such as data center virtualization and vertical markets. The key challenge for these companies is to offer enough value in areas customers care about. With hardware prices plummeting and enterprises more security-conscious than ever, harvesting spare cycles to save money is a harder sell than it used to be.

The environments described below are different. The grids manage the servers, networks and storage devices that deliver core business services, rather than stealing spare capacity from machines doing something else. That may give them a better chance of success in an environment where everyone is circling the wagons to focus on mission-critical activities.

Resource management

Resource management is the core element of grid computing. Grids must be able to manage different kinds of hardware and software to make them appear as a single unit, with appropriate provisioning, security and monitoring functions. When one part of the grid is a Sun server and another a cluster of Linux PCs, for example, ensuring that processes automatically fail over to backups if an element goes down is tricky. But there are real efficiencies to be gained in abstracting out resources. "If you buy a new refrigerator, you don't call the local power company and tell them you need another kilowatt of power," notes Rich Friedrich, director of the Internet systems and storage lab at HP.

Like Web services, grid resources are usually loosely coupled, meaning that no one has absolute control over their constituent pieces. But grids and Web services address different issues. Web services operates at the level of applications, with standard protocols to federate or transmit data and mechanisms to define component-based applications. Web services focus on what you want to do. Grids, by contrast, operate at the lower levels of the stack. They address the myriad issues of how to get to a world of distributed applications and service-oriented architectures.

From Science to Commerce, By Way of Web Services

Academic roots

The most prominent grid computing efforts began in academia, including the Globus Project, led by Ian Foster of Argonne National Laboratory and the University of Chicago, Steve Tuecke of Argonne, and Carl Kesselman of the Information Sciences Institute at the University of Southern California. (Coincidentally, ISI was the home of the late Jon Postel, caretaker of the Internet domain name system.) Globus is an open-source codebase and a set of de facto protocols for implementing grids. The Globus project began in 1996 and is now widely used for scientific high-performance computing initiatives.

In recent years, the scientific grid community has run smack into the developing world of commercial distributed computing technologies, including peer-to-peer (P2P), clustering and Web services. The result has been a fruitful collaboration. Emerging today is a new synthesis, sometimes called grid services, drawing elements from both worlds. "Grid systems can be reasonably complex. We need industry contributing to the development of the technology," says Foster of the Globus Project.

The turning point came last year, when IBM made a major commitment to grid computing. In addition to funding internal development work, IBM joined with Globus to develop an architecture for merging grids with Web services. It has brought its significant resources to bear as an active promoter of the grid concept. IBM sells plenty of equipment for scientific high-performance computing, but the company linked its grid involvement specifically to its commercial potential. Irving Wladawsky-Berger, vp of the server group, has identified grid computing as the third in a chain of technologies IBM took advantage of that emerged from the scientific community, the others being Internet-centric e-business and Linux.

Grids meet Web services

As the IT industry rushes headlong into the brave new world of Web services (SEE **RELEASE 1.0**, SEPTEMBER AND OCTOBER 2001), it is finding that protocols alone don't make systems run efficiently, handle exceptions, scale in response to demand or distribute themselves to the appropriate hardware. Grids turn out to be a perfect complement for where the commercial sector has been going. Foster elaborates: "I'm interested in Web services because it provides us with a convenient platform for delivering the technologies we've been working on for a while. We're looking for a delivery vehicle, and business is looking for a management function."

Though companies such as Grand Central and Flamenco Networks, described in part II of our Web services report, as well as Blue Titan (the former ServiceMesh) (SEE **RELEASE 1.0**, MARCH 2002), are going after aspects of the problem, they don't have the deep focus on distributed resource management that is the focus of much grid computing work. "If we believe in service-oriented architectures, the networks themselves should be built as sets of services," explains Glass of The Mind Electric.

Standards and politics

The official framework for the merger of grids and Web services is the Open Grid Services Architecture (OGSA). Originally a project of IBM and Globus, OGSA is now being developed under the auspices of the Global Grid Forum (GGF), a standards body originally formed from the merger of American, European and Asian scientific grid computing research organizations. GGF now includes over 200 organizations, with major sponsors including IBM, Microsoft, HP, Sun, Qwest, Platform Computing and Silicon Graphics.

Two articles co-authored by Foster, "The Anatomy of the Grid" and "The Physiology of the Grid" (SEE RESOURCES SECTION) are the most comprehensive descriptions of the concepts of grid computing and its relationship to Web services. OGSA is a broad framework for grid-connected Web services; the core technology components are known as open grid services infrastructure (OGSI). Working groups with GGF are refining standard interfaces and behaviors of grid services under OGSI, as well as ancillary elements such as security.

In the Web services world, standards efforts have degenerated thanks to a profusion of overlapping standards and sniping between traditional rivals Microsoft and Sun. Fortunately, GGF has emerged as something like the IETF of distributed computing – the place where everyone comes together to hash out standards. Two other standards efforts, the New Productivity Initiative led by Platform Computing (SEE PAGE 11) and the Intel-founded Peer to Peer Working Group, both merged into GGF.

There is some concern that GGF is too closely tied to IBM, which contributed significant funding to the group, or to Globus, which is not the only major academic grid computing platform. And Sun has held up completion of some OGSA standards by raising questions about patents. However, IBM announced recently that would license its contributions to OGSI on a royalty-free basis. A breakdown is still possible, but so far OGSA is the only game in town, with strong pressure for standards.

Big Gridders

IBM, HP, and Sun all have recently launched significant grid computing initiatives, just as they are major players in Web services (SEE CHART BELOW).

Microsoft hasn't traditionally been a significant player in the high-performance computing or data center market. It doesn't have a significant grid initiative today, though it agreed in March to fund a Windows XP port of the Globus Toolkit, and its .Net effort to create a distributed software development and delivery environment is entirely consistent with grid developments. Microsoft also has developed intellectual property around distributed resource management in its labs and through its Universal Plug and Play initiative for connecting devices. Depending on what Microsoft chooses to do, the 2005 launch of Longhorn, the next version of Windows, could be the most significant grid computing development of all.

IBM: utility computing über Alles

IBM has put the most muscle behind grids. It supports leading scientific grid efforts, include the National Science Foundation's distributed terrascale facility (DTF). DTF will be the world's most powerful grid, with an interconnected series of Linux clusters delivering 13.6 teraflops of computation and storing more than 600 terabytes of data. (Sun is also involved in the project.) IBM has related initiatives, autonomic computing and e.Liza, to build self-healing and self-organizing devices.

	Web services initiative	Grid initiatives	Comments
IBM	WebSphere	Autonomic; E.Liza; IBM Grid Computing	Biggest push, but relationship among elements unclear
HP	HP Web Services	Utility Data Center; Planetary Computing	Data center management focus
Sun	Sun ONE	N1; Grid Engine	Broad, phased vision
Microsoft	.Net	N/A (though porting Globus to Windows)	Could become a major player

FIGURE 1: *Summary of major vendor grid-related activity.*

Earlier this month, new ceo Sam Palmisano made these projects the centerpieces of e-business on demand, IBM's vision for utility computing. Today's IBM is a service-centric company, so it's only natural that it would promote a service-oriented computing model. IBM's major grid customers today are primarily in the academic, government, and research realms, including the US Department of Energy, the British government, and University of Pennsylvania.

IBM is positioning its Global Services unit as a grid systems integrator, and is making its Tivoli systems-management software and WebSphere application server environment grid-aware. Like most recent-vintage IBM activities, the effort is a mix of in-house software and hardware, new technology from IBM research and partner products. It's trying to cover both the high-gloss vision and the mundane execution. IBM research chief Paul Horn's autonomic computing manifesto (SEE RESOURCES) is a lyrical and compelling argument for adopting biological approaches to IT management, while IBM grid computing gm Tom Hawk emphasizes more prosaic benefits: "One of the true promises of grid computing is application integration."

HP: reinventing the data center

HP's grid efforts draw heavily on its recent acquisition of Compaq, which following its own spate of acquisitions was a leading high-performance computing vendor. The combined company makes about half of the world's 500 most powerful supercomputers, and it sees those machines evolving into "power centers" for the emerging computing grid.

HP's grid initiative, Utility Data Center, is focused on the data center market, though it also has a long-term program in HP Labs called Planetary Scale Computing that has a broader mission of providing universal access to global computing resources. HP credits itself with a more solution-oriented focus than its competitors, which it sees exploiting grids largely for marketing value. Rather than focus on individual devices, HP attempts to enhance the data center as a whole. Says HP's Rich Friedrich, "Grids can begin to address the operational costs of enterprise computing, and in most enterprises, that's the biggest cost."

Sun: N1 as the platform layer

Sun recently announced an initiative called N1 to deliver a suite of grid-oriented products. "The definition of a service has changed. Instead of being one piece of software on one box, it's running on a lot of pieces of hardware. But it's still one service," explains Adam Hawley, senior marketing manager for N1.

Sun already has a successful resource management product, Grid Engine, with customers including Ford, Motorola, Ericsson and BP. It offers a free version for Solaris and Linux which supports over 6,000 grids. In the past, it has focused more on clustering together its proprietary high-end machines than on opening up possibilities for commodity computing. Where IBM and HP tout their broad product lines and

large services organizations, Sun has always focused on integrated solutions that play well with others but take advantage of its proprietary technologies.

N1 is both more incremental and more radical than competitors' initiatives. Its product roadmap starts with a virtualization engine that supports various platforms including Solaris, NT and Linux and doesn't require significant re-architecting of data centers. This allows customers to aggregate pool of resources in their data centers, similar to other grid solutions. In late November, Sun bough data center automation startup Terraspring to incorporate its technology into N1.

Going forward, Sun plans to offer higher levels of abstraction on this foundation, including service provisioning and policy-oriented automation. In effect, it's building a true network-scale operating system for data centers. Companies will identify business needs, and the network will arrange itself to support those needs by shifting around resources and creating virtual services.

Sun sees N1 complementing its Sun ONE strategy for Web services. Says Yousef Khalidi, chief architect of N1, "Strategically we see Sun ONE as the application layer and N1 as the platform layer. And they do meet."

Grid Specialists

The leading grid startups all have roots in academic and research work on distributed computing. They generally look to complement, rather than compete with, the IBM/HP/Sun grid onslaught, by focusing on narrower markets or offering greater depth for particular functions.

The distributed computing startups we covered in our P2P report – Popular Power, United Devices, and Data Synapse – could also be considered grid players. Popular Power failed to find funding and shut down, but the other two remain in business, as does competitor Entropia (SEE BOX, PAGE 14). However, they are more oriented toward cycle sharing than towards the deeper resource management function that we see as central to grids. The companies below may be hitting the market at a better time (earlier in Platform's case; later for the others).

Platform Computing: ten years and going strong

Platform, the largest and most established grid vendor, has been selling clustering and resource management software for ten years and now has 1500 customers, which puts it somewhere between the startups and the IT giants. Its traditional strength is in high-performance computing and in markets such as electronic design automation for chip development and product design in enterprises.

PLATFORM COMPUTING INFO
Headquarters: Markham, Ontario, Canada
Founded: 1992
Revenues: $46 million in 2001
Employees: 400
Funding: organic from sales
URL: www.platform.com

"We're still in the same old game," says founder and cto Songnian Zhou. "The new buzzword is grid computing, but from my perspective, it's just large-scale distributed computing." Zhou developed the resource allocation algorithms that formed the basis of Platform's flagship LSF software during his PhD research at UC Berkeley and serving as an intern at HP in the late 1980s. His combination of enthusiasm, rapid-fire delivery and an accent make him difficult to follow at times, but Zhou's vision for computing has proven right on target.

When he started Platform, Zhou says, he talked with IBM, HP and Sun, but none of them appreciated the need for distributed workload management. They weren't concerned with lower server utilization because they made money selling more servers. Zhou found this approach short-sighted: "You don't have a house, a backup house and a test house. When your relatives come, they know where to sleep!"

Now, says Zhou, things have changed. Vendors are pushing grids, and it's customers who don't know what to make of the concept: "People look around and ask, 'What is the killer app of grids?' It's all around you, in pieces! Nothing has to be brand new. It's all evolution. If we can take these pieces and link them together, using whatever protocol we use, over five to ten years the infrastructure shifts to utility computing."

As applications become more network-aware and performance demands increase, the only viable strategy is to scale out as well as up, Zhou explains, an approach he labels "10,000 ants instead of one elephant." The new enterprise IT stack he envisions has distributed resource management software at the base, supporting application servers and Web servers, which in turn support the applications and, increasingly, services.

Platform is supporting IBM's grid and Web services efforts, and Zhou serves on IBM's e-business council. "The key marriage is Web services with grids," Zhou notes. "Web services are the application infrastructure, building on app servers. Grids are the systems infrastructure supporting applications." Platform also offers agent-based

instrumentation and monitoring tools to help customers measure utilization and manage their grids more effectively. Using standard business intelligence tools such as Cognos, customers can model and test the required infrastructure for a service, and plan for demand efficiently rather than over-provisioning.

Platform offers a variety of software packages for different aspects of distributed computing, from simple clusters up to complex grids, including a customized version of the Globus toolkit. It recently released Platform Symphony, a workload orchestration and service provisioning engine focused on the financial services market. JP Morgan Chase is using it to aggregate several thousand servers to deliver Internet-based applications.

The challenges now for Platform are the familiar ones of execution and navigating the relationships with larger platform vendors. Its 150 employees in R&D are more than the total staffs of any of the startups below, though those companies think they can leapfrog Platform by building in grids and Web services from the start. Zhou says he is seeing a combination of top-down implementations, where cios dictate that they will implement grids and the software architects who work for them have no idea where to start, and bottom-up situations where a departmental clustering effort grows into an enterprise grid.

Avaki: our advantages are legion

Avaki, formerly known as Applied MetaComputing, traces its roots to the Legion project at the University of Virginia. Legion is the second most prominent academic grid computing environment, with more of an integrated vision and architecture compared to the Globus approach of offering a suite of lower-level resource management services. Andrew Grimshaw, who led the Legion Project, is the founder and cto of Avaki. The company first developed a demonstration project to simulate protein folding in conjunction with the Scripps Research Institute. Its software became commercially available earlier this year.

AVAKI INFO
Headquarters: Cambridge, MA
Founded: August 1999
Employees: 40
Funding: $16 million from Polaris Venture Partners
URL: www.avaki.com

Avaki purchased the Legion source code, licensed it back to the University of Virginia for research purposes, and has spent the past 20 months enhancing the software for commercial applications. Like Legion, Avaki focuses on providing secure, transparent access to resources, so that applications can run across collections of computers seamlessly. Many companies, says president Tim Yeaton, are struggling with data currency and access challenges, and Web ser-

vices alone won't help. Web services provides a development and delivery model to integrate disparate systems, but no way to marshal data.

Avaki has a strong focus on the collaborative aspects of grids. Grimshaw believes it's important to see grid computing as more than just sharing processing capacity. "Grids are really about resource sharing in distributed, possibly multi-organizational settings," he says. "Those resources could be computing resources, but they can also be data resources." And CPU cycles turn out to be the cheapest and easiest resources to share. The real business challenges turn on higher-level structures.

For example, a group of biotech companies collaborating on drug development will want to share their data in a way that is secure and protects their intellectual property. They need a way to create a virtual computer between the organizations that provides access to just the appropriate information. That requires an infrastructure to identify and manage resources across different, potentially incompatible systems. (We described a related set of issues in our report on collaboration infrastructure. SEE **RELEASE 1.0**, FEBRUARY 2002.) "Grids are the way to facilitate sharing as narrow as the business relationships between the companies define," says Yeaton.

Awkward basic aggregation of computing cycles by parceling out jobs, such as the SETI@Home project to search space for signals from alien civilizations, are the bottom rung of the distributed computing hierarchy, Grimshaw thinks. Next come clusters and load-management systems such as Platform's LSF and Sun's Grid Engine. Avaki focuses on environments with higher-level data coherence and latency requirements, which require sophisticated federation and caching. Its grid software usually complements existing clustering and load-management systems.

Distributed identity

Identity and security are also critical for these applications. "If I don't have a name for something in another place, I don't have a way to refer to it. That's a fundamental property of what differentiates a grid from something else," says Grimshaw.

Avaki's software provides its own naming infrastructure, consistent with emerging OGSA standards, allowing any resource on a grid to have a cryptographically secure identifier. As a result, Avaki can support grids that work across organizations without compromising security and data consistency. "You want to do naming and security at the beginning of the grid, and not add it later, which is a problem with the Internet – not putting these in from the beginning," Grimshaw points out.

ENTROPIA: GRIDS ON THE DESKTOP

While most grid companies today are focused on data centers and servers, Entropia still sees an opportunity in aggregating spare processing capacity of desktop PCs. The company has survived where most desktop distributed processing startups failed, in part because of its technical street cred. Its cto, Andrew Chien, is a computer science professor at UC San Diego and a leading grid expert. Entropia's scientific advisory board includes Globus leaders Ian Foster and Carl Kesselman.

Entropia, founded in 1997, initially offered software that ran across the public Internet, but it now concentrates on working within enterprises. "We are focused on providing a platform for customers to run any native Win32 application across a network of PCs," explains svp of product development Bill Philbin.

Running a grid across end-user PCs poses unique challenges. The machines have lower-bandwidth and less-reliable connectivity than servers in a data center, and the grid application can't interfere with whatever the person using the PC is doing. Entropia's technology ensures that end-user performance and security aren't affected by distributed processing in the background. "Our definition of unobtrusiveness is that when things go bad, the job should silently quit. We should find ways to do that job somewhere else," explains Philbin.

Philbin acknowledges that some tasks, such as transaction processing and applications requiring shared memory, are ill-suited for PC grids. In other cases, though, desktop processing can complement and extend customer's high-performance infrastructure. A grid of 1000 desktop PCs, Philbin points out, would offer the processing power of one of the world's 500 most powerful supercomputers. Entropia's customers are in similar fields to the data-center grid vendors: financial services for risk-management and analytic applications; life sciences for genomics and related activities; and media for high-end graphics rendering.

Today, when companies wish to work together and share information, they face unattractive choices. They could give each other accounts on their networks, but that exposes much more than the specific resources they want to share. The alternative, creating new infrastructure outside both companies' firewalls, involves significant costs and human effort. Avaki lets the participants set access controls to expose only the information they wish to share with the other company.

On a technical level, Avaki implements a distributed trust model by building access controls into objects themselves. "From the very beginning we believed it was absolutely imperative to couple identity in the computer science sense of the word with a name," explains Grimshaw. Each name has embedded cryptographic keys, allowing any two objects to mutually authenticate themselves and perform authorized actions. Avaki has proposed a specification, secure grid naming protocol, to the Global Grid Forum in an effort to establish standard naming mechanisms across grid implementations.

Avaki also distinguishes itself from competitors by offering its grid software as a runtime rather than a toolkit, meaning that it can be installed out of the box without custom coding. Customers can have a grid up and running in a matter of hours, though typical deployments take a few days to weeks for custom integration. The company has several paying customers in life sciences, including Pfizer, Gene Logic, NC BioGrid, Structural Bioinformatics, and Texas Tech University. In addition to

supporting OGSA, Avaki is making its software fully compatible with industry standards such as J2EE, so that it fits better into existing corporate IT environments.

Because of its unique identity-oriented approach, Avaki complements many other grid players; it has partnerships with IBM, HP and Sun, the three large vendors that could be seen as competitors. Avaki is an active participant in OGSA/OGSI working groups, and hopes to be one of the first commercial vendors to support those standards as they are ratified next year.

The Mind Electric: bringing the grid to life

The Mind Electric (TME) is the clearest example of a company that sees a direct line from Web services to grid computing. The company's first product, Glue, is a lightweight environment for developing Web services that has found favor among developers for its simplicity. But while competitors such as Systinet and CapeClear (SEE **RELEASE 1.0**, SEPTEMBER-OCTOBER 2001) focus on broadening their Web services offerings, TME's second offering is designed explicitly to enable grid environments.

"As things become more and more wired together, there's going to be all kinds of changes in the way people architect and manage systems," founder and ceo Graham Glass explains. The first phase of Web services has catalyzed interest in service-oriented architectures and experimentation with component-based software. "Once enterprises start doing that," Glass continues, "they are going to try to figure out how to assemble those services into robust scalable systems."

That requires new infrastructure to handle functions such as failover, resource management and security in a heterogeneous, distributed environment. Glass thinks the timing is now right, with 2003 as the big year for grid takeup. "Any large enterprise in America that we encounter building Web services is trying to figure out how to control, manage and audit them," Glass concludes.

THE MIND ELECTRIC INFO
Headquarters: Addison, TX
Founded: February 2001
Employees: <20
Funding: organic and from angels
URL: www.themindelectric.com

Glass, who has a background in distributed computing research, previously founded distributed application toolkit vendor ObjectSpace, which sold off its products in early 2002 to Recursion Software. He has a reputation for visionary pronouncements about the parallels between computing and biological complexity. The name of the company points in this direction, as does that of the new product: Gaia. It's a reference to an ancient earth goddess and the hypothesis of biologists Lyn Margulis and James Lovelock that our planet is a self-regulating living system.

Glass recognizes that today's market wants working solutions rather than compelling ideas. Though Gaia was always the centerpiece of the company's plans, it chose to develop Glue first in order to get marketplace traction and developer acceptance. If TME had a reputation for easy-to-use development tools, Glass figured, it would be simpler to convince customers to try out the more far-reaching distributed computing concepts Gaia invokes. Glue has more than 50 customers, including AT&T, Countrywide Mortgage, Bank of America, Fujitsu, GE Global Exchange, Millennium Pharmaceuticals, TRW and Vodafone. These are the best prospects for Gaia, which be available early in 2003 for free downloading in beta form.

It's services all the way down

Glass describes Gaia as "a runtime grid services platform which itself is organized as a collection of services." It transparently offers the registration, discovery, failover, management and monitoring features necessary to take enterprise Web services to the next level. Using Gaia, a developer who wants to publish or subscribe to a grid service adds a few lines of Java. Neither the client nor the server runs the actual Gaia code; it sits on intermediary servers that self-assemble into a cloud.

Behind the scenes, Gaia broadcasts the service's availability throughout the network, distributes work among nodes, and handles failover if a node goes down. When a new client joins the grid, it sends out a ping, finds a neighbor, and uses a P2P search algorithm to find other nodes and available services. "The grid is providing an awful lot of very rich infrastructure invisibly, in the cloud," says Glass. The system works with major development environments such as .Net, enterprise Java beans and Perl.

Asked to distinguish grid computing from current enterprise IT infrastructure, Glass resorts to the ever-popular electricity analogy. Application servers such as BEA WebLogic or IBM WebSphere provide quality of service for the Web servers they support directly, but not for other computing resources on the network. They function like self-contained generators. The grid layer, like the electric power grid, allows administrators to manage performance uniformly for a service or collection of services, rather than at each server.

Glass sees TME having a competitive advantage in being independent of complex legacy technologies, in which category he includes such things as EJB. "We don't have any political reason not to do what we think will be most loved by developers," he argues. And though TME is avoiding grand pronouncements that might scare off

chastened cios, Glass hasn't abandoned his ideals. "Our hope is that by releasing Gaia, we'll demonstrate to enterprises that this can be fun and exciting," he says.

Making the Grid

Virtually there

Grid computing represents the clearest and most comprehensive manifestation of the distributed model for IT infrastructure. Technology has been moving in this direction, both in research settings and in production environments, for many years. Now, spurred by scaling and cost-containment demands, distributed computing is becoming a focal point the IT industry. Many localized developments are moving in the same direction.

Storage vendors are pushing virtualization, though which pools of disk drives can be treated interchangeably. Service providers, software vendors and integrators focused on the data center, including Opsware (formerly LoudCloud), Think Dynamics, AppStream, VMWare and Polyserve are making it easier to provision servers without manual configuration of individual boxes when capacity needs change. And every major enterprise hardware vendor is coming out with blade servers, designed from the ground up to be rack-mounted and managed in groups. Without low-level resource management, though, higher-level abstractions can go only so far.

Power to the people

Some applications simply require more computing power or other resources than a single device can economically deliver. In the academic realms where grid computing developed, this necessity was the mother of invention. Fields such as particle physics, fluid dynamics, cryptography and weather forecasting involved massive data sets and processing requirements, often with limited budgets. Grids were the only way to harness the necessary resources.

Today, academic researchers are not the only ones who need massive computing power and care about budgets. Financial-services firms engage in highly quantitative portfolio analysis and derivatives risk modeling, and pharmaceutical companies develop drugs though complex modeling of genes and proteins. In a down economy, these companies can't keep throwing hardware at these problems; they need to make better use of the servers and storage devices they already have.

Moreover, as companies focus more on services and work to make their offerings more personalized and realtime (SEE **RELEASE 1.0**, DECEMBER 2001) the need for massive computational power on demand increases. A trader may be able to run a Monte Carlo simulation in batch mode overnight, but when an investment bank wants to give its best customers personalized and immediate portfolio analysis, that's not good enough. Demands are particularly intense in regulated industries.

Complexity and decentralization

There is more and more complexity inside the typical enterprise data center. One consequence of the move from mainframes to client-server to n-tier Web-based e-business is a dramatic increase in complexity. No longer does one vendor control all aspects of the infrastructure within a company. The proliferation of intranets, extranets, and departmental enterprise applications means that important computing resources are no longer safely controlled within the glass house of MIS; they are spread throughout the organization.

"One of the largest complaints or challenges that IT organizations share with me is that the environment they operate in is too complex," says IBM's Tom Hawk. "In some cases, IT is now being viewed as an inhibitor to business growth." The decentralization of enterprise IT is a good thing. However, it has come at the expense of cost, performance and reliability, three characteristics that are now front and center for most businesses. No one wants to go back to the inflexible, closed environment of the 1960s mainframe. But there is great demand for tools that allow for better server utilization and more effective service provisioning and monitoring.

The evolution of grids was in some ways inevitable. Over the long haul, industries based on computers are shaped by the relative improvement of three characteristics: processing, storage, and bandwidth. They all improve over time, but at different rates. In recent years, bandwidth and storage prices have fallen even faster than processing, which follows the well-known path of Moore's Law. Where processing power doubles every 18 months, bandwidth (measured in terms of maximum transmission capacity, not the phantom demand growth rates that WorldCom and others trumpeted during the boom) has doubled roughly every 12 months. Storage, the ugly duckling in this equation, has grown fastest of all, with capacity doubling roughly every nine months.

Put these trend-lines together, and here's what you get. Instead of building more computers, it's efficient to network together the computers we already have, "wast-

CONSUMER GRIDS

Grids aren't just for buttoned-down business users. An important sector for the high-performance computing and clustering market is entertainment. Rendering for digital animation, for example, requires huge server farms running jobs in parallel.

The demise of Napster and most of the P2P startups in its wake cut off short-term possibilities for Internet-based end-user grids. Yet one of Napster's file-sharing successors, Kazaa, has embedded in it a distributed computing application that harvests unused processing from users' PCs for network-wide applications. Discovery of this "Altnet" capability provoked some outcry, because most users weren't aware of it... despite the language buried in the end-user license agreements they had to click on to download Kazaa.

While grids using end-user PCs are problematic, grids may serve consumers in other ways. Sony's Playstation 3 gaming console, scheduled for a 2005 release, will use a novel CPU developed with IBM and Toshiba, called Cell. This chip will have multiple processing cores, allowing it to shift between tasks such as computationally intensive 3D graphics and high-speed networking.

In another consumer-oriented example, The Mind Electric has a joint venture with Hitachi to apply grid computing to wireless handheld devices. The software allows the devices to find each other, establish ad hoc networks and share information without tying back to local wireless hotspots. Hitachi has a license to deploy the technology throughout Japan.

But the most promising consumer grid application is massively multiplayer online games (SEE **RELEASE 1.0**, OCTOBER 2002). Microsoft, Nintendo and Sony are all launching online gaming for their consoles, and PC-based role-playing games such as Ultima Online and Everquest are doing bang-up business.

Now a startup based wants to apply grid computing to make such games even more powerful. It's called Butterfly.net. And it's headquartered in West Virginia, not exactly the center of the technology world, but a place where office space is cheap and rural development tax incentives and grants are plentiful.

"Today, in the games, the main problem is reliability," says Butterfly ceo David Levine. Everquest has half a million members, and is supported by a large number of servers. The entire game goes down for 10 to 12 hours several times a month to allow for patches to those servers, frustrating gamers. When many players are in the same physical location in the virtual world, performance can grind to a halt, because each server has limited capacity. Moreover, the online worlds must be segmented, because resources can't run across the entire network of servers.

Butterfly has built a special-purpose grid, starting with the Globus Toolkit, for online gaming. Every computer in the grid, as well as every end-user device, runs stripped-down physics models of all activities in the virtual world that could potentially affect their users. Central game servers control the authoritative model of the world. They can determine and make adjustments when nodes are out of phase with one another. This way, says Levine, "A player can move from game server to game server just like you move from city block to city block. You have the illusion of gameplay spanning the servers."

The company has several game developers as customers, though most of them don't want their names released. Levine says Butterfly is currently managing seven development projects, which he hopes will eventually move to deployment.

Butterfly has signed an agreement with IBM to host and manage the actual grid of machines on which Butterfly's software runs. Levine says he's even gotten calls from cios of universities interested in becoming customers, because they realize their students are playing online games and they want someone else to manage the bandwidth required without themselves losing the ability to share in the revenue the games generate.

The big difference between online gaming and most of the applications described in this issue is the need for low latency. Levine explains the difference colorfully: "Say you're in battle, everyone is following the arrows into the air... and you don't know what happens for six hours. You might as well be playing chess!" Butterfly's software must anticipate where capacity will be needed, so that it can shift resources instantly. In the future, Levine suggests, the technology could be useful in low-latency industrial settings as well. Areas such as process control, logistics, and dynamically priced transactions could represent significant opportunities.

Butterfly.net has 12 employees and has raised $2.3 million from Walker Ventures, Adena Ventures and the West Virginia Jobs Investment Trust.

ing" bandwidth to economize on relatively scarce processing power. And it's efficient to cache multiple copies of data on relatively cheap storage systems, thereby saving both bandwidth and processing.

Grids and service business models

Though the term comes from the electric power industry, grid computing doesn't require an on-demand utility business model. That model may emerge for some markets, such as online gaming, but most grid computing will take place within organizations or webs of partners. Grids may enable new services and they may make IT infrastructures more effective and efficient, but they won't necessarily be visible to end-users.

Predictably, this will cause some teeth-gnashing. Because of how heavily vendors such as IBM are promoting grids, and the association with P2P cycle harvesting over the public Internet, there is an expectation that massive public grids are right around the corner. When it becomes clear they aren't, grids will be attacked as another piece of industry hype. The real work and real implementations, however, will continue beneath the radar. Grid computing techniques will eventually be used to support on-demand services, but only when the business conditions are right. The difficulties of back-end integration for distributed systems were one reason so many application service providers (ASPs) failed. "ASPs weren't really achieving the economies of scale they needed to make the business work, because the virtualization technology wasn't in place," says Foster.

COMING SOON

- Digital identity for objects.
- The service grid.
- Human capital.

- And much more. . . (If you know of any good examples of the categories listed above, please let us know.)

It's easy being grid

Grids are still early in their implementation cycle for commercial applications. Many deployments today are local "intragrids" on a single network. As products become more robust and standards more widely implemented, company-wide and cross-corporate grids will become more common.

Grids may even have an impact on demand for bandwidth. Grids lend themselves to high-powered applications and services that span large geographic areas. Many of those, including visualization (a field we discussed last month), particle physics experiments and online gaming, involve significant bandwidth requirements. For example, starting in 2005 the large hadron collider at CERN will

Resources & Contact Information

Tim Yeaton, Andrew Grimshaw, **Avaki**, 1 (617) 374-2500; fax, 1 (617) 374-2520; tyeaton@avaki.com, agrimshaw@avaki.com
David Levine, **Butterfly.net**, 1 (304) 260-9537; fax, 1 (304) 260-9529; david.levine@butterfly.net
Bill Philbin, **Entropia**, 1 (858) 623-2840; fax, 1 (858) 623-2847; bphilbin@entropia.com
Ian Foster, **Globus Project**, 1 (630) 252-4619; fax, 1 (630) 252-5986; foster@mcs.anl.gov
Rich Friedrich, **HP**, 1 (650) 857-5024; fax, 1 (650) 236-1612; rich_friedrich@hp.com
Nick Vanderzweep, **HP**, 1 (408) 447-3436; fax, 1 (408) 447-7113; nick_vanderzweep@hp.com
Tom Hawk, **IBM**, thawk@us.ibm.com
Graham Glass, **The Mind Electric**, 1 (972) 764-5115; fax, 1 (972) 764-3215; graham@themindelectric.com
Owen Davis, **Petal Computing**, 1 (917) 721-5572; owen.davis@sonatainc.com
Songnian Zhou, **Platform Computing**, 1 (905) 948-4200; zhou@platform.com
Adam Hawley, Yousef Khalidi, **Sun Microsystems**, 1 (650) 960-1300; adam.hawley@sun.com, yousef.khalidi@sun.com

For further reading:

Ian Foster, Carl Kesselman, Jeffrey Nick and Steven Tuecke, "The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration," (2002), available at http://www.globus.org/research/papers/ogsa.pdf
Ian Foster, Carl Kesselman and Steven Tuecke, "The Anatomy of the Grid: Enabling Scalable Virtual Organizations," (2001), available at http://www.globus.org/research/papers/anatomy.pdf
Global Grid Forum – http://www.gridforum.org
Autonomic Computing: IBM's Perspective on the State of Information Technology – http://www.research.ibm.com/autonomic/manifesto

generate petabytes of data per year through its experiments, which will be analyzed through GriPhyN, the grid physics network. We may see an uptick in backbone capacity requirements from such computer-to-computer applications before the impacts of human-driven uses such as residential broadband are felt.

Says Foster of the Globus Project, who has been doing grid computing for years, "I think right now we're just working on the edges of where the real benefits are." ■ R 1.0

The Fairmont Scottsdale Princess
March 23 to 25, 2003

Who? What? Where? Data Comes Alive

As always, the issues addressed at PC Forum (like the industry itself) are continuing to take shape. Several months out, here's our focus today:

How will we use all the capacity - hardware, communications and software - built up during the bubble? We're going to use it not to build more castles in the air, but to get a better grasp of the world that IT systems are supposed to be tracking, analyzing, organizing and optimizing. Users, corporate and individual, want their IT systems to answer some basic questions:

Who? means identity management, which underlies security and access systems, provisioning, customization, personalization, and CRM. All systems will be role-based, designed to recognize their users - employees, customers, partners, prospects - and to respond appropriately.

We'll be hearing from two leaders in this field: **Oblix** ceo **Gordon Eubanks** and **Critical Path** ceo **Bill McGlashan**. But identity management is relevant to all IT systems. It's key to such companies as **Amazon.com** and **AOL** with their personalization features, as long promoted by invited speaker **Jeff Bezos** and as about to be better leveraged by speaker **Jon Miller**. It's also key to finding terrorists, as you'll hear from **Jeff Jonas** of **SRD**, which got its start tracking questionable individuals in Las Vegas.

What? means digital ids for objects. Systems must be designed for better tracking of supply chains and for managing inventories of products, spare parts, designs and work in process (whether tangible or virtual). The government wants to track things for security; companies want to lower costs and improve efficiency. As systems start to talk to one another, the data standards companies use internally must cede to ways of representing things that operate across corporate and national boundaries. Data interoperability is as important as software interoperability.

Some industries are farther ahead in such standardization than others. Leaders include companies such as Wal-Mart; we'll be hearing from its former cio and current **Sam's Club** ceo **Kevin Turner**. Identifying and tracking objects is also key for Amazon.com and for every company that owns real-world assets. Of course, keeping track of data is what speaker **Larry Ellison**'s **Oracle** has done from the very beginning.

Where? means mapping and location. For consumers, the big "location" applications are driving directions and online travel services. But soon all digitally identified objects will be locatable in space. Logistics, scheduling, marketing and most real-world processes depend on knowing where people and things actually are and how they are spatially related. Location-based services will also be a big driver behind mobile phone applications. As speaker **Mark Cattini** of **MapInfo** will discuss, the Internet has moved beyond the stage of making location irrelevant; now we can use it to leverage location.

We'll do all this with interoperable software - Web services. We'll cover the promise and the reality

of Web services with a panel moderated by John Hagel, author of next month's *Release 1.0* and of the new book *Out of the Box* (both about Web services) with **Jerry Hale** of **Eastman Chemical** and **John McKinley** of **Merrill Lynch**. **Salesforce.com**'s **Marc Benioff** will lay out the advantages and challenges of the ASP delivery model.

As for why: The motivation comes from people, not from systems. In the end, as vendors, we want all this information to make things different in the real world: more efficient, more relevant, more pleasing to humans. As users, we want what we were promised all along: information, efficiency, awareness, insight. We'll be getting that from the speakers listed below and from you!

The value of the Forum is not just in the formal sessions, but in the conversations provoked among you and your fellow attendees - like-minded and intelligently different-minded industry leaders, corporate customers, investors, analysts and journalists. Come improve your mind as well as your fortunes at PC Forum!

PC Forum 2003 Speakers

Confirmed

Marc Benioff, *Chairman & CEO*, SalesForce
Sergey Brin, *President, Technology & Co-founder*, Google
Robert Carter, *EVP & CIO, Co-CEO, Fedex Services*, Fedex
Mark Cattini, *President & CEO*, MapInfo
Gordon Eubanks, *President & CEO*, Oblix
Larry Ellison, *Chairman & CEO*, Oracle
John Hagel, *Author*
Jeff Jonas, *Chief Scientist & Founder*, SRD
Bill McGlashan, *Chairman & CEO*, Critcal Path
John McKinley, *EVP & Head of Global Technology & Services*, Merrill Lynch
Jonathan Miller, *Chairman & CEO*, AOL
Paul Otellini, *President & Chief Operating Officer*, Intel
John Seely Brown, *VP & Chief Scientist*, Xerox

Invited

Peka Ala-Pietila, *President*, Nokia
Jeff Bezos, *Chairman, President & CEO*, Amazon
Michael Capellas, *Chairman & CEO*, WorldCom
Gilman Louie, *CEO*, In-Q-Tel
Jeff Raikes, *Group Vice President*, Microsoft
Dan Rosensweig, *Chief Operating Officer*, Yahoo!
Jonathan Schwartz, *Executive VP, Software*, Sun Microsystems
Martin Sorrell, *Group Chief Executive*, WPP
Kevin Turner, *President, CEO & Executive VP*, Sam's Club/Wal-Mart

To register and for speaker updates, please visit
http://www.edventure.com/pcforum/

The conversation never stops! Subscribe to our free email newsletter, *The conversation continues,* for thought-provoking analysis from our editors, along with commentary from our highly intelligent readers. Sign up at **http://www.edventure.com/conversation/join.cfm**.

Release 1.0 Subscription Form

Complete this form and join the other industry executives who regularly rely on *Release 1.0* to stay ahead of the headlines. Or if you wish, you can also subscribe online at www.release1-0.com.

Your annual *Release 1.0* subscription costs **$795** per year (**$850** outside the US, Canada and Mexico), and includes both the print and electronic versions of 11 monthly issues; 25% off the cover price when you order from our online archives; a *Release 1.0* binder; the bound transcript of this year's PC Forum (a $300 value) and an invitation to next year's PC Forum.

NAME _____

TITLE _____ COMPANY _____

ADDRESS _____

CITY _____ STATE _____ ZIP _____ COUNTRY _____

TELEPHONE _____ FAX _____

EMAIL* _____ URL _____

**personal email address required for electronic access.*

☐ My colleagues should read Release 1.0, too!
Send me information about multiple copy subscriptions and electronic site licenses.

☐ Check enclosed ☐ Charge my (circle one): AMERICAN EXPRESS MASTER CARD VISA

CARD NUMBER _____ EXPIRATION DATE _____

NAME AND BILLING ADDRESS _____

SIGNATURE _____

Please fax this form to Natasha Felshman at 1 (212) 924-0240.

Payment must be included with this form. Your satisfaction is guaranteed or your money back.

If you wish to pay by check, please mail this form with payment to: EDventure Holdings, 104 Fifth Avenue, 20th Floor, New York, NY 10011, USA. If you have any questions, please call us at 1 (212) 924-8800; email us@edventure.com; www.edventure.com.